

## DOCUMENT RESUME

ED 191 069

CS 205 790

AUTHOR Alloway, Evans  
TITLE Holistic Scoring Session for the New Jersey Writing Project: A Summary.  
SPONS AGENCY New Jersey State Dept. of Education, Trenton.  
PUB DATE May 79  
NOTE 23p.: Report prepared by the New Jersey Writing Project. Several pages may be marginally legible.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.  
DESCRIPTORS \*Evaluation Criteria; \*Grading; Minimum Competency Testing; Secondary Education; Student Writing Models; \*Testing; \*Writing (Composition); \*Writing Skills  
IDENTIFIERS \*Holistic Scoring; New Jersey; \*Writing Evaluation

## ABSTRACT

The holistic procedure used to score 2,800 student papers in the New Jersey Writing Project is described in this booklet. Characteristics are given of papers that received high, middle, and low scores. It is noted that (1) papers receiving high scores were characterized by a clear understanding of and a direct approach to the assignment, a well-presented opinion, a clear focus, an appropriate organizational pattern, a sense of closure, a sense of style, and relative freedom from errors; (2) papers receiving middle scores were characterized by inconsistency; and (3) papers receiving low scores contained a significantly greater number of simple sentences, lacked focus, and were often seriously flawed grammatically. A sample essay at each level is provided. Attachments include a summary of how to score essays holistically and the actual assignment used in the New Jersey Writing Project. (MKM)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED191069

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
NATIONAL INSTITUTE OF  
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

HOLISTIC SCORING SESSION  
FOR THE  
NEW JERSEY WRITING PROJECT:

- A SUMMARY

by Evans Alloway  
Educational Testing Service  
May 1979

New Jersey Writing Project  
One Executive Drive  
Monmouth Junction, New Jersey 08852

\* \* \*

This booklet was prepared under a grant funded  
by New Jersey State Compensatory Education  
Research and Development.

"PERMISSION TO REPRODUCE THIS  
MATERIAL IN MICROFICHE ONLY  
HAS BEEN GRANTED BY

New Jersey Writing  
Project

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC).

FEB 26 1980

2

CS205790

Scores awarded to 2,800 student papers provided one portion of the evaluation of the New Jersey Writing Project. Prior to receiving instruction, students in both treatment and control classes wrote papers in response to a single assignment; that is, all students wrote on the same assignment. Near the end of the semester, in one-semester classes, and near the close of the year in full-year classes, those students wrote again in response to the same assignment. The papers were then coded and mixed so that the readers would not know either who had written a paper or whether the writing had occurred before or after instruction. The aim of the scoring session was to rank all the papers on the basis of their overall quality and then to answer two questions: 1) Did papers written after instruction receive scores significantly higher than those awarded to papers written before instruction? 2) If they did, is the difference between pre-instruction scores and post-instruction scores greater for papers from treatment or control classes? Discussion of the results appears elsewhere in the report of the evaluation of the New Jersey Writing Project.

To score the papers, the Project employed twenty-four teachers--twenty as Readers and four as Table Leaders. The writer of this summary served as Chief Reader and thus directed the selection of the sample papers that served as models for the scorers, oriented the scorers to the task and to the sample papers, and resolved discrepant scores.

The scorers read the papers holistically. ETS has used holistic scoring to evaluate students' writing for many years, the practice now serving as the basis of scoring in almost every program that asks the test candidates to provide samples of their writing. Among those programs are the California State University and College System English Placement Test, the New Jersey Basic Skills Assessment Test, the State Department's test for Foreign Service Officer, and College Board tests in the College-Level Examination Program, Advanced Placement Program and Admissions Testing Program. Attachment A, excerpted from a document written for the teachers who score papers for the Admissions Testing Program's English Composition Test, presents a full summary of the rationale and procedures for holistic scoring; similar procedures guided the scoring session for the New Jersey Writing Project.

4

The scoring session provided two independent judgments for each paper, each judgment recorded as a score on the four-point scale, 1-2-3-4. For purposes of reporting, the judgments for each paper were summed, providing a seven-point reporting scale, 2-3-4-5-6-7-8.

It is clear, of course, that scorers rendered judgments on a comparative basis. That is, a paper received a top score if it was judged to be among the best papers in the batch; a paper received the lowest score if it was judged to be among the least good papers in the batch. Even though scorers awarded scores on the basis of comparison, some comments about the papers themselves are nevertheless in order.

First, the assignment, which appears as Attachment B, worked well in that it stimulated the students to write and it enabled every student to say something; there were very few blank papers. The assignment did present some ambiguities, however. For one, the students defined "object" much more broadly than did the examiners. Many students, possibly more engaged by the discussion of the object's importance or origin than by the object itself, chose to define "object" very broadly indeed. Thus, whereas the examiners had in mind objects like a penknife, a musical instrument, a lucky charm,

or an article of clothing, many students wrote about life, a friend or relative, a class trip, or a favorite activity. Faced with a similar situation in the classroom, the teacher-examiner can acknowledge and discuss the ambiguity with the students and, perhaps, ask them to write again as necessary. Not so in an assessment situation in which only the examiners can deal only with what the students put before them. Keeping in mind that the evaluation was designed to look at writing skill and not the ability to read the examiners' minds, the leaders of the scoring session decided to accept for scoring any paper that attempted to be responsive to the assignment. For example, if a paper had as its subject, say, a class trip but had as its focus the importance of the trip to the writer, the paper was judged scorable. Readers eased their classroom consciences by defining "object" as "something" and then defining "something" as "nearly anything."

The invitation to the students to write in any one of several modes provided yet another challenge to the readers. Because readers must compare writing performances with other writing performances, an evaluation normally asks writers to prepare papers in one mode only. The leaders of the New Jersey Writing Project felt a great concern for the kind of statement that the

evaluation would make about writing, however, and they chose to retain the invitation in order to confirm that writing consists of more than expository prose. The fact remains, however, that readers can only opine about what effect the choice of mode has on a writer's score.

As for the writing itself, papers that received high scores were obviously different from papers receiving lower scores. High-scoring papers, for one thing, were marked by a clear understanding of and direct approach to the assignment. Their writers presented an opinion and supported it well, providing a clear focus, an appropriate organizational pattern, and a sense of closure. These papers were generally consistent, demonstrated a sense of style, and, though perhaps flawed to some extent, were relatively free of surface errors. The following paper represents to some extent papers receiving comparatively high scores. The reader is reminded, however, that there is no such thing as "the" top paper. So long as they share a level of quality, papers of many sorts can receive the same score. (See attached sample essay KK.)

Papers receiving middle scores were characterized by, among other things, inconsistency. Generally competent, they nevertheless lacked the level of linguistic sophistication demonstrated

by the papers ranked above them. Middle papers presented some variety of sentence patterns, but it is significant that real problems often occurred when the student writer attempted to break away from the familiar subject-verb-predicate pattern. The following sample paper may present some indication of the level of quality among middle papers, but again the reader is reminded that other papers receiving the same score may present other characteristics. (See attached sample essay S.)

Papers receiving the lowest scores were often seriously flawed. Clausal expressions rarely deviated from the subject-verb-complement pattern, and low papers contained a significantly greater number of simple sentences. Many of these papers also lacked focus, some of them providing ample discussion but without demonstrating an awareness of the limits of the assignment. Low papers lacked mastery of the conventions of standard written English and, while spelling was generally not a deciding factor so far as scores were concerned, they contained more misspellings than papers ranked above them.

It is significant that most papers written by students for whom English is not the first language find themselves in the lower portions of the scoring scale. The following paper is typical of many papers receiving low scores. (See attached sample essay P.)

#### Writing Sample KK

I have succeeded in extending my view of the world through a rather simple object -- a camera. I have made permanent thousands of fleeting images, captured time for a second and preserved it in silver on a thin strip of plastic. I own a time machine.

I go nowhere without it, not chancing ~~that~~ the loss of a crucial instant which ~~will be lost~~ I might encounter without it. It is my visual memory, more accurate and revealing than the one God gave me. It neither embellishes nor detracts and it never forgets what it sees.

In my darkroom I have stored my images: friends and relatives frozen forever, frames from the movie of life. Most don't even know I have these pictures and would be embarrassed to see them. I cherish them all, especially the ones everyone else dislikes. The posed shots show the least, the candid the most.

My camera preserves what I cannot. It is interesting to consider that my images will last for longer than I will, until people are forgotten and become merely interesting curios. Until even my camera breaks down, leaving only its memories, lasting until ~~it~~ they are useless: faithfully preserving what ~~it~~ they saw but having no one to interpret ~~its~~ their views.

My camera truly does act as a time machine. Currently it allows me to step back a few days or years. Soon it will carry me forward, past my mortal life, transporting my image into a future I shall never know.

### Writing Sample S

The objects that are most important to me or were most important to me are a series of buildings known to all as Merkins Poultry Farms.

These buildings have special meanings to me because it was there that I had my first real paying job. The buildings themselves are pretty raunchy looking and smelling but I feel that through them and the year that I worked with them I matured a great deal. At first I was the original servile worker who obeyed the bosses wishes to the "T". I needed them more than they needed me. The bosses themselves were slave driving people who would never you a break and would work you till you dropped. They treated their workers like one of the 60,000 hens that they owned. But you always hoped that deep down they were basically good human beings. The incoming money for me was great. For the first time I had spending money but after a while they just shunted me aside and didn't call me for work again for 3 months until Easter when they needed me again. After Easter was over and things were back to normal they pushed me aside again even though I had, over Easter, done all the rotten jobs nobody else wanted to do. At this point I would have begged for work which now I realize how stupid it was. When summer came they decided that I was "good enough" to work for them. I soon became their very best worker. I don't consider this bragging but the truth. I would work harder, I could pack the eggs faster and do almost every job better than anyone on the farm. At this point (after the summer was over) they practically begged me to work for them, which I agreed to only out of asinine loyalty to people who never commended by work.

I finally quit this past weekend after realizing I had out grown my very first job. I had grown enough to realize my pride was more important then the dollar.

## Writing Sample P

### Life

I discovered that my life was important in life from when I was young to now. That I was born into this world because I was wanted and I discovered that I rembered stuff about myself from the age of 2. The reason I guess was that I didn't have a full memory for alittle while and when I did I remembered. Anyway life has had its up's and down's turnaround's etc. for me. If it didn't I would have been lying because a life has to have faults in it along the way. It's all apart of growing up. If life wasn't meant to happen than when this world was created they wouldn't have had people on this earth if they didn't want them. Then it just took its natural course from there on. It all depends on the person and how they want to run their life and other on his etc. to find out the different kinds of ~~live~~ lifes and ways people behave about themselves. It's self discipline and alot of other things too.

Reading the Essay Holistically

Each paper is read by at least two readers. The essays are read quickly for an over-all impression. Readers are asked to react immediately to a paper and to score it on the basis of that first impression. They are asked not to reread papers or to analyze specific points of composition.

The basic assumptions of the holistic reading are that each of the factors involved in writing skill is related to all the others and that no one factor can be separated from the others. Readers must judge each essay as a whole; they must read each paper for the impression its totality makes. A misspelled word, a comma splice, a sentence fragment, a misplaced modifier should carry no great weight in scoring a paper. The candidate is entitled to make some mistakes; he or she is writing hurriedly in a tense situation, without recourse to a dictionary, without the customary time for deliberation. If readers read each paper as a whole, then, they are better able to judge the competence of the writer. If the paper is poorly written, that will be part of the reader's first impression; there is no need to analyze it word by word to decide that it is badly done. If the paper contains mistakes because the writer, though capable, has been forced to hurry, the reader will

judge the general quality of the paper most effectively by reading it as a whole. The motto of the reading might well be "Read quickly and judge. Do not reread."

Of course, holding rigidly to that motto would be ridiculous. There are some papers a reader cannot judge immediately. Such papers should not be read and reread, but rather taken to the table leader immediately. The table leaders are there to insure that each paper is read as nearly as possible in the same way. Whenever a reader cannot make a quick judgment, he or she must unhesitatingly ask the table leader for help.

#### The Conduct of the Reading

The organization of the reading process is a simple one. To maintain the quality of the reading, individual readers are responsible to a table leader; each table leader is responsible to one of the three or four assistant chief readers; each assistant chief reader is in turn responsible to the chief reader. All of these people work together to set the standards for the reading and to maintain them. The reading is very definitely a cooperative effort. The table leader is the person the reader should consult when he or she has questions about any paper. If necessary, the table leader will call upon the help of the assistant chief reader. Certain essays, such as those that are off the topic,

illegible, or obscene, should automatically be given to the table leader. So should essays that are really poems, short stories, or plays. These are special papers; they are not to be dismissed lightly, but should be scored according to the scoring policy set for them.

Throughout the reading, the reader's work is checked constantly by the table leader. It is essential to have all papers read the same way. Perfect agreement among large numbers of readers is not possible, but we must act as if it is and work toward achieving it. Everything possible is done to maintain the scoring standards, and in such an effort the cooperation of all readers is vital.

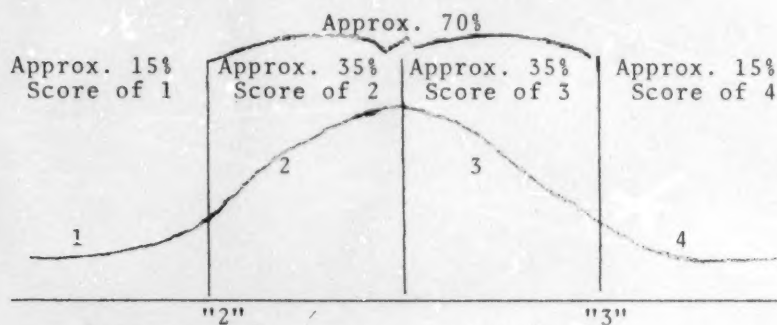
The procedures during the reading and the setting of the standards for the reading follow a pattern:

1. The chief reader presents the topic to the readers and attempts to define the writing problem inherent in the topic. He or she answers certain basic questions about the scoring of the papers: How much attention should be paid to misspelling? What is considered off the topic? What do we define as a specific example?
2. The chief reader explains the scoring scale to be used. The four scores range from 4

(the highest) to 1 (the lowest). In a scale which has only four points (4, 3, 2, 1), there is obviously no middle score. One of the major reasons why the scale works as well as it does, is that readers are forced away from a middle, or uncommitted, score. Readers must first decide whether the paper is above average (in the 3-4 category) or below average (in the 1-2 category). After making this first decision, readers must then judge whether the paper is enough above average to rate the highest score (4) or enough below average to rate the lowest score (1). Often some new readers are hesitant about granting any papers a 4. They need not be. Among the thousands of papers produced by candidates throughout the country there are certainly many that are enough above average to receive a 4, just as there will be papers that deserve 3's. Unfortunately, no reader--experienced or inexperienced--seems to need assurance about giving out 2's and 1's; what all readers seem to need from time to time is the reminder that not all the papers are "2" papers or "1" papers. Though the fatigue of reading may make it seem so, not every candidate who writes an essay is below

average in writing ability, and the readers must constantly guard against giving scores much lower than the papers deserve.

The scores for a large population such as the one taking the College Board Achievement Tests usually follow a normal curve like the one below.



For our purposes, the line through the highest point of the curve represents a score of 2.5--the middle score that individual readers cannot give. On either side of the middle line are lines representing (to the left) a score of 2 and (to the right) a score of 3. In a so-called normal distribution, almost seventy percent of the scores should fall between the lines for 2 and 3. The rest should fall evenly on either side of the 2 and 3 lines; that is, should have scores of 1 and 4.

In other words, a reader would, if he or she were to follow an exact "normal" distribution, award as many 4's as 1's, as many 3's as 2's. He or she would not award more 1's than 3's, for example. However, we do not ask a reader to award a nearly perfect distribution of 15% 4's, 35% 3's, 35% 2's, 15% 1's. In fact, readers are asked not to keep a tally of the scores they have given. We simply ask them to bear in mind at all times that some candidates are above average in ability and deserve to be scored as such. Not only should there be above average papers but, because the group is being judged against itself, there must be such papers. The candidate's performance is not being measured against an ideal but against the performance of all the other candidates writing an essay on this topic for this particular test.

3. After the discussion of the rating scale and of an ideal distribution of scores, the chief reader presents a series of sample papers. This group of papers has been judged previously by the chief reader and the assistant chief readers as representative of particular categories of competence.

paper, a just below-average paper, a poor paper, and perhaps one or two others.

Readers are asked to rank the papers.

When the readers have completed that task, the chief reader announces the score that the table leaders, the assistant chief readers, and the chief reader have given each paper. It is up to each reader to compare the way he or she ranked the papers with the way they were ranked by experienced readers.

(All sample papers used to instruct readers are duplicates of essays written by candidates taking the test. Immediately after the administration of the test the chief reader and his or her assistants read hundreds of papers to determine how students have performed on this test on this topic. They then select samples representative of the range of quality as it is found. They do not score the essays against an ideal standard, but against the standard established by the actual performance of students writing on this particular topic.)

4. The chief reader then distributes to readers another set of sample papers, one at a time. After each reader has scored the sample papers, the chief reader asks for a show of

hands from the readers to indicate the scores given to each paper. The scores given by all the readers are tallied and reported to the group. At that point, the chief reader announces the score that the chief reader, assistant chief readers, and the table leaders thought the paper deserved. When there is a great difference between this score and the score the readers have given, the chief reader will explain the reason the experienced group gave that particular score. Those reasons may or may not be accepted by the group of readers. The chief reader does not prescribe the scores; he or she offers guidelines, in the light of experience. It is the consensus of the readers themselves that determines the standards.

Always, during this process, the individual reader is responsible for noting in what category the particular score awarded the paper fell. Is the reader scoring with the majority of the readers? If not, is he or she scoring too high? Too low? It is necessary for all readers to adjust to the standards of the majority. There can be no exceptions if all papers are to be scored

as nearly as possible in the same way. An individual reader who assumes that he or she has the right to set individual standards is being unfair to the candidates whose papers he or she will read.

As sample papers are read and their scores recorded, some papers will receive from all the readers an almost equal number of, say, 2's and 3's. This division in scoring should alarm no one, for obviously there are papers that are exactly average and would be rated as 2.5 if such a score were possible, just as there are papers that deserve 3.5 or 1.5--other scores that individual readers cannot give. However, the rating system is designed to provide for such scores.

Each paper is read two times, and each reader gives that paper a score ranging from 1 to 4. Therefore, when these scores are summed, a paper may receive a score as low as 2 or as high as 8. Because not every paper receives the same score from each reader, a paper may have a final score of 2, 4, 6, 7, 8, or any other score on the scale. If one reader judges a paper as 3 and another as 4, no one is greatly alarmed. We expect such differences occasionally.

However, we do not expect readers to give 1's to papers that other readers have scored 4. That large a discrepancy in scoring cannot ordinarily be tolerated. However, very few papers under very special circumstances may seem to deserve such divergent scores.

3. The process of scoring samples, of reporting scores by a show of hands, and of tallying these scores continues until the chief reader is sure that the readers are in close agreement, that they have a clear understanding of the bases for judging papers, and that, insofar as is humanly possible, all papers will be scored in the same way. The chief reader then announces that the actual scoring of the papers is to begin, even though the table leaders may continue to work individually with a few readers who are still not scoring with the group. In order to ensure reliable results, these readers' scores must, of course, be brought into line with those of the other readers.
6. From time to time throughout the reading, readers will be given other sample papers to read and score. This continuous reading of samples, as well as the numerous rest periods, helps readers keep to the standards

they have set. After all, when readers have read a few hundred essays on the same topic, their fatigue naturally causes them to feel that subsequent papers lack a certain freshness. Scores do tend to drop after the first two days of the reading. Educational Testing Service does all that it can to see that this very human reaction does not result in unfairness to individual candidates. For example, the papers are distributed among readers in a way that ensures that each paper will have its first scoring during those first two days. During the reading the readers and the chief reader must also do what they can to see that this drop, which is only to be expected, does not seriously damage the quality of the scoring.\*

\*From How the Essay in the College Board English Composition is Scored. Gertrude Conlan. ETS, 1978.

## ATTACHMENT B

### WRITING TOPIC

Write on this topic for twenty minutes.

Write about an object that is really important to you, something that has become a part of your life. You might want to write about how you discovered it or what it has come to mean to you. You may write a journal entry, a letter, a story, a brief autobiography, an essay, or a poem.